

**SYSTEM AND METHOD FOR ROUTER MULTICAST CONTROL**

Thomas C. McDermott III  
265 Daniel Drive  
Plano, Texas 75094  
Citizenship: USA

Jim Kleiner  
13834 Sprucewood Drive  
Dallas, Texas 75240  
Citizenship: USA

**RELATED APPLICATIONS**

This application is related to co-pending and commonly assigned U.S. Application Serial Number [59182-P001US-10020638], entitled "System And Method For IP Router With an Optical Core," to co-pending and commonly assigned U.S. Application Serial Number [59182-P002US-10020639], entitled "System and Method for Router Central Arbitration," to co-pending and commonly assigned U.S. Application Serial Number [59182-P004US-10020641], entitled "System and Method for Router Data Aggregation and Delivery," to co-pending and commonly assigned U.S. Application Serial Number [59182-P006US-10020643], entitled "Timing and Synchronization for an IP Router Using an Optical Switch," to co-pending and commonly assigned U.S. Application Serial Number [59182-P012US-10021641], entitled "Router Network Protection Using Multiple Facility Interfaces," to co-pending and commonly assigned U.S. Application Serial Number [59182-P013US-10021642], entitled "Router Line Card Protection Using One-for-N Redundancy" and to co-pending and commonly assigned U.S. Application Serial Number [59182-P014US-10021643], entitled "Router Switch Fabric Protection Using Forward Error Correction," all filed October 31, 2000, the disclosures of which are incorporated herein by reference.

# TECHNICAL FIELD

This application relates to the field of optical communication networks, and particularly to large-scale routers for optical communication networks.

0914426 111800

## BACKGROUND

Routers with electronic switch fabrics have the ability to store data packets and then make multiple copies of those packets. In one approach, for example, a packet is copied into an electronic fabric and each output card reads the same copy of that packet. When all of the outputs have read the packet, it can be discarded. Thus the fabric itself is capable of performing a multicast operation, in which one input packet is cloned and copied to multiple outputs.

A router switch fabric that has no inherent storage or replication capability, for example an optical switch fabric, cannot fundamentally multicast packets. There are other approaches that, for example, store a packet on an input line card in electronic form and have the input line card make multiple copies of the packet, sending them to each multicast output as necessary. Although this will function and perform the multicast operation, a drawback is that the packet sits in memory on the input card, and if sufficient numbers of multicast packets arrive, the memory on the input card can be exhausted, or the memory can back up subsequent nonmulticast packets and delay or block them from getting through the switching fabric.

## SUMMARY OF THE INVENTION

The present invention is directed to a system and method to transfer a multicast packet out of an input card which may have unicast as well as multicast traffic coming into it, and to remove the multicast packet to a dedicated multicast card. The dedicated multicast card is then responsible for replicating multiple copies of that packet to the various multicast outputs. This approach is applicable to router switching fabrics, for example optical switch fabrics, that have no inherent packet storage or replication capability. A dedicated multicast card is substantially the same kind of card as an input and output line card, but without external facility interface connections.

In an embodiment of the present invention an incoming multicast packet is transferred through the router switching fabric to a dedicated output multicast card substituting for an ordinary output line card, which then converts the packet from an optical packet to an electrical packet and passes it to a dedicated input multicast card substituting for a conventional input line card. There the packet is stored and can be replicated electrically, converted back to an optical packet, and transferred through the router switching fabric to multiple destination output line cards. In a preferred embodiment, the input and output dedicated multicast cards are in reality the same dedicated multicast card.

Transfer of the multicast packet to multiple destinations can occur sequentially over multiple switching cycles or can occur simultaneously. In some embodiments the switching fabric is subdivided into multiple subplanes, such that multiple parallel switching paths exist through the switching fabric. This enables simultaneous parallel multicasting to multiple destinations during a single switching cycle. Some embodiments of the present invention include multiple dedicated multicast cards. This configuration allows rapid expansion of the multicast tree to a large number of destinations.

Various aspects of the invention are described in co-pending and commonly assigned U.S. Application Serial Number [59182-P001US-10020638], entitled "System And Method For IP Router With an Optical Core," co-pending and commonly assigned U.S. Application Serial Number [59182-P002US-10020639], entitled "System and Method for Router Central

Arbitration,” co-pending and commonly assigned U.S. Application Serial Number [59182-P004US-10020641], entitled “System and Method for Router Data Aggregation and Delivery,” co-pending and commonly assigned U.S. Application Serial Number [59182-P006US-10020643], entitled “Timing and Synchronization for an IP Router Using an Optical Switch,” co-pending and commonly assigned U.S. Application Serial Number [59182-P012US-10021641], entitled “Router Network Protection Using Multiple Facility Interfaces,” co-pending and commonly assigned U.S. Application Serial Number [59182-P013US-10021642], entitled “Router Line Card Protection Using One-for-N Redundancy” and co-pending and commonly assigned U.S. Application Serial Number [59182-P014US-10021643], entitled “Router Switch Fabric Protection Using Forward Error Correction,” all filed October 31, 2000, the disclosures of which are incorporated herein by reference.

The foregoing has outlined rather broadly the features and technical advantages of the present invention in order that the detailed description of the invention that follows may be better understood. Additional features and advantages of the invention will be described hereinafter which form the subject of the claims of the invention. It should be appreciated by those skilled in the art that the conception and specific embodiment disclosed may be readily utilized as a basis for modifying or designing other structures for carrying out the same purposes of the present invention. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope of the invention as set forth in the appended claims. The novel features which are believed to be characteristic of the invention, both as to its organization and method of operation, together with further objects and advantages will be better understood from the following description when considered in connection with the accompanying figures. It is to be expressly understood, however, that each of the figures is provided for the purpose of illustration and description only and is not intended as a definition of the limits of the present invention.

## BRIEF DESCRIPTION OF THE DRAWING

For a more complete understanding of the present invention, reference is now made to the following descriptions taken in conjunction with the accompanying drawing, in which:

Fig. 1 is a block diagram illustrating multicast data packet flow through a router system having an optical switching fabric, in accordance with an embodiment of the present invention; and

Fig. 2 is a block diagram illustrating a router embodiment that includes multiple dedicated multicast cards.

5  
0097788.1

## DETAILED DESCRIPTION

Fig. 1 is a block diagram illustrating multicast data packet flow through a router system 10 having an optical switching fabric 12, in accordance with an embodiment of the present invention. Optical switching fabric 12 can be partitioned into multiple switching subplanes, for example subplanes 12-1 through 12-3, each having multiple inputs and outputs, which are interconnected dynamically in input/output pair configurations through the optical switch, for example through switching paths 106 and 112-1 through 112-3. The inputs of optical switching fabric 12 are interconnected with multiple input line cards, represented for clarity in Fig. 1 by input line card 14 interconnected through optical link 105 with optical switching fabric 12. Each input line card 14 is in turn interconnected through an electrical link 103 with an input facility interface card 13. Input facility interface card 13 is connected through optical input data path 101 with other network systems, for example peer routers. The outputs of optical switching fabric 12 are interconnected with multiple output line cards, represented for clarity by output line cards 20-1 through 20-N interconnected with optical switching fabric 12 through optical links 113-1 through 113-N. Output line cards 20-1 through 20-N are further interconnected through electrical links 114-1 through 114-N with respective output facility interface cards 21-1 through 21-N, which in turn are connected through respective optical output data paths 115-1 through 115-N with other systems across the network.

In the present embodiment, a dedicated output multicast card 16 is interconnected through an optical link 107 with an output of optical switch fabric 12. Likewise, a dedicated input multicast card 18 is interconnected through an optical link 110-1 with an input of optical switch fabric 12. Dedicated multicast cards 16 and 18 are substantially the same type of cards as respective input line card 14 and output line cards 20-1 through 20-N, but without external facility interface connections. Dedicated multicast cards 16 and 18 are connected together through a data loop 108.

When a packet destined to be multicast (a "multicast packet") arrives at an input line card 14, the packet header destination is decoded and the multicast packet is accordingly

forwarded through optical links 105 and 107 and through optical switch fabric 12 along switching path 106 to dedicated output multicast card 16. Dedicated output multicast card 16 converts the optical multicast packet to an electrical multicast packet, which is looped back to dedicated input multicast card 18 through data loop 108. Dedicated input multicast card 18 then converts the multicast packet back to an optical multicast packet, which it forwards through optical link 110-1 and along a switching path 112-1 to an output line card 20-1 interconnected with a destination optical switch output through an optical link 113-1. At output line card 20-1 the multicast packet is handled in the same fashion as any other data packet arriving at output line card 20-1.

A system and method for a network router with an optical switching fabric is described in U.S. Application Serial Number [59182-P001US-10020638], cited above. Although for the purpose of clarity corresponding input and output functions are shown on separate circuit cards in separate  $\frac{1}{2}$  line shelves, embodiments are described in which corresponding input and output functions are combined on a single circuit card in a single line shelf combining  $\frac{1}{2}$  input and output line shelves, thus creating a folded router system configuration. Particularly, input and output packet forwarding line cards can be combined into a single physical circuit card in a single line shelf. In this folded system configuration, dedicated output multicast card 16 and dedicated input multicast card 18 are in fact the same circuit card, and data loop 108 can become simply a conductive trace on that single circuit card.

Once the multicast packet is removed from the original input line card 14, it no longer sits in memory on the input line card, and thus no longer delays or blocks ordinary network traffic. In some embodiments dedicated multicast card 16, 18 makes multiple replicas of a multicast packet and then forwards these replicas through optical switch fabric 12 to multiple switch destinations connected with multiple output line cards 20-1 through 20-N. These replica multicast packets can be forwarded serially during separate switch cycles.

Alternatively, replica multicast packets can be forwarded simultaneously within one switch cycle to multiple destinations, provided that optical switch fabric 12 incorporates m



multiple switching subplanes, for example subplanes 12-1 through 12-3, that support  $m$  multiple parallel switching paths, illustrated schematically by switching paths 112-1 through 112-3. In the router system described in U.S. Application Serial Number [59182-P001US-10020638], cited above, for example, a multiple subplane switch fabric architecture supports simultaneous parallel switching paths. This structure allows sending two or more multicast replica packets in parallel simultaneously from one dedicated input multicast card 18 through switch fabric 12 to multiple destination output line cards 20-1 through 20- $N$  along separate paths, for example a path through link 110-1 and then through switching path 112-1 and a separate path through link 110-3 and then through switching path 112-3. If the required multicast output size were six, for example, one multicast packet could be sent from input line card 14 to a dedicated multicast card 16, 18, which could replicate the packet and transmit up to three multicast replicas in parallel to three destination output line cards, for example output line cards 20-1 through 20-3, during one switch cycle, and then during a subsequent switch cycle transmit another three multicast replicas in parallel to three additional destination output line cards, for example output line cards 20-4 through 20-6. This multicast operation requires three switch cycles, namely one cycle to send the original multicast packet from the input line card to the dedicated multicast card, and two additional cycles to transmit multicast replicas to the six destinations at three replicas per cycle. Because the first cycle is pipelined, the actual delay is two cycles.

The number of times  $t$  that a multicast packet is sent through the switch fabric is given by an expression

$$t = 1 + \text{CEILING}\left(\frac{N}{m}\right)$$

one time to go from the input line card to the dedicated multicast card and then

$\text{CEILING}\left(\frac{N}{m}\right)$  times, where  $N$  is the number of destination line cards,  $m$  is the

number of simultaneous parallel switching paths, and the *CEILING* function rounds up a value to the next larger integer.

Fig. 2 is a block diagram illustrating a router embodiment 50 that includes  $M$  multiple dedicated multicast cards 16-1 through 16- $M$  and 18-1 through 18- $M$ . Switch fabric 12 is partitioned into  $m$  multiple subplanes, for example subplanes 12-1 through 12-3, thus supporting  $m$  multiple simultaneous switching paths, for example switching paths 106-1 through 106- $m$ , 112-1 through 112- $m$ , and 116-1 through 116- $m$ . In a folded router geometry, multicast cards 16-1 and 18-1, and connecting data loop 108-1 occupy a single circuit card. Similarly multicast cards 16- $M$  and 18- $M$  and connecting data loop 108- $M$  occupy a single circuit card. In the example of Fig. 2, a multicast packet is sent by input line card 14 simultaneously through parallel switching paths 106-1 through 106- $m$  across switch fabric 12 to  $M$  dedicated multicast cards 16-1, 18-1 through 16- $M$ , 18- $M$ , where the packet is replicated at each multicast card and then sent in  $m$  parallel replica packets from each dedicated multicast card 16-1, 18-1 through 16- $M$ , 18- $M$  through simultaneous switching paths 112-1 through 112- $m$  and 116-1 through 116- $m$  respectively across switch fabric 12 to  $N$  destination line cards 20-1 through 20- $N$ . This multiple replication provides an  $N$ -to-one multicast tree expansion. If two dedicated multicast cards 16- $M$ , 18- $M$  can each send three replica packets, then three replicas multiplied by two multicast cards would provide six-to-one multicast tree expansion. This procedure improves the fan out and reduces the time delay, since more data is multicast within a single switch cycle.

In the more general case with  $N$  multiple output destinations,  $M$  multiple dedicated multicast cards, and  $m$  simultaneous parallel switching paths, the number of times  $t$  that a multicast packet goes through the switch fabric is given by an expression

$$t = \text{CEILING}\left(\frac{M}{m}\right) + \text{CEILING}\left(\frac{N}{mM}\right),$$

where variables and functions are as defined above. If for example  $M = 2$ ,  $m = 3$ , and  $N = 100$ , then  $t = 18$  switch cycles, whereas if  $M = 5$ ,  $m = 3$ , and  $N = 100$ , then  $t = 9$  switch cycles, which is only half of the previously calculated delay. The  $m$  parallel switching paths limited by the subplane structure of the switch fabric refer only to those paths that originate

from a common input, for example either a common input line card or a common dedicated multicast input card. The  $M$  different multiple dedicated multicast cards can be increased independent of this limitation. The  $M$  different multiple dedicated multicast cards can then each originate a set of  $m$  simultaneous parallel switching paths through the switch fabric.

5 Since adding a dedicated multicast card is a lower cost option than adding a switch subplane, multiple dedicated multicast cards provide an advantageous approach to increasing the multicast tree expansion with minimum switch cycle delay.

This means that for example in Fig. 2, one input line card 14 could send a multicast packet simultaneously to  $M$  different dedicated multicast cards, which then could perform multiple replication. Thus, the number of multicast cards times the number of available simultaneous parallel switching paths expands the multicast tree for a single switch cycle. This affords the flexibility to install as many multicast cards as is necessary to support the required multicast tree expansion, with an option to install fewer if desired.

Although the present invention and its advantages have been described in detail, it should be understood that various changes, substitutions and alterations can be made herein without departing from the spirit and scope of the invention as defined by the appended claims. Moreover, the scope of the present application is not intended to be limited to the particular embodiments of the process, machine, manufacture, composition of matter, means, methods and steps described in the specification. As one of ordinary skill in the art will  
20 readily appreciate from the disclosure of the present invention, processes, machines, manufacture, compositions of matter, means, methods, or steps, presently existing or later to be developed that perform substantially the same function or achieve substantially the same result as the corresponding embodiments described herein may be utilized according to the present invention. Accordingly, the appended claims are intended to include within their  
25 scope such processes, machines, manufacture, compositions of matter, means, methods, or steps.